PSYCHOACOUSTICAL SIGNAL PROCESSING FOR THREE-DIMENSIONAL SONIFICATION

Tim Ziemer

University of Bremen
Bremen Spatial Cognition Center
Medical Image Computing
Enrique-Schmidt-Str. 5, D-28359 Bremen
ziemer@uni-bremen.de

Holger Schultheis

University of Bremen
Bremen Spatial Cognition Center
Institute for Artificial Intelligence
Enrique-Schmidt-Str. 5, D-28359 Bremen
schulth@uni-bremen.de

ABSTRACT

Physical attributes of sound interact perceptually, which makes it challenging to present a large amount of information simultaneously via sonification, without confusing the user. This paper presents the theory and implementation of a psychoacoustic signal processing approach for three-dimensional sonification. The direction and distance along the dimensions are presented via multiple perceptually orthogonal sound attributes in one auditory stream. Further auditory streams represent additional elements, like axes and ticks. This paper describes the mathematical and psychoacoustical foundations and discusses the three-dimensional sonification for a guidance task. Formulas, graphics and demo videos are provided. To facilitate use at virtually all places the approach is mono-compatible and even works on budget loudspeakers.

1. INTRODUCTION

Just like visual displays, auditory displays can serve for various applications in numerous scenarios. Many ubiquitous auditory displays are information-poor alerts and alarms, like the doorbell, horn, siren and alarm clock [1, 2, 3]. Here, the information is binary (on/off). Other auditory displays carry more information, like the Geiger counter, which sonifies radiation on a ratio scale [4] (i.e., a continuous scale with a natural zero). Even more information is sonified in pulse-oximetry [5, 3], where heart rate is presented on a ratio scale and, simultaneously, oxygen concentration on an interval scale (i.e., a continuous scale without a natural zwro). [4] point out that developers of auditory displays have to make sure that relations in the data are heard correctly and confidently by the user. Likewise, [1] state that the perceived information should match the intended message. At the same time researchers face issues when trying to add further information to an auditory display. They experience that orthogonal, i.e., independent, acoustical parameters perceptually interact [6, 7, 8].

Some researchers avoid this issue by leveraging spatial audio. The human listener is able to localize sound sources in three-dimensional space. Hence, sonification for navigation in one-[9, 10], two-[11], and three-dimensional space[12] often employs spatial audio by means of binaural headphone presentation or loudspeaker arrays. Authors report intuitive and successful use, especially in combination with visual guidance. However, the highest localization precision of audible sources is $1 \pm 3°$ along the azimuth in the front [13, 14]. It becomes worse by one order of magnitude towards the sides and in the median plane. Distance estimation has a resolution of decimeters in the near surrounding and degrades drastically with increasing distance. Listeners can distinguish some dozens locations in the horizontal plane, a little less in the median plane and along the distance dimension. For many applications, this spatial resolution is not sufficient. Binaural presentation and sound field synthesis methods further degrade localization precision and may cause additional localization phenomena, like in-head localization, front-back confusion, a vague distance perception, elevation and, sometimes, azimuth errors[15, 16]. To improve sonification in three-dimensional space, [12] added monaural cues as redundant elevation and distance cues.

Other authors suggest mapping multidimensional data completely to monaural sound attributes. The approach is promising as we can distinguish for example between 640 and $4,000$ pitches [17, ch. 7][18, p. 136], 120 loudness steps [17, ch. 7] and 250 sharpness steps (cf. [17, ch. 9] and [19]). [4] realized the need of a set of orthogonal parameters that adequately span hearing perception. [1] recognize that implementing psychoacoustical modeling and synthesis is a challenging task. It is an inverse problem because perceptual sound attributes cannot be controlled directly via signal processing. Only physical sound attributes can be manipulated. If anything, the perceptual outcome of physical parameter magnitudes can be predicted. [7] argue that psychoacoustic models provide no implementable guidelines to achieve this, because they are only valid for certain, mostly static, test signals. Still, [6] formulate two suggestions to solve the problem. The first is to create massive lookup tables to solve the inverse problem by looking up physical audio parameter magnitudes that create the desired magnitudes of all perceptual attributes. The downside of this approach is that continuous, subtle changes in desired sound attributes are created by discontinuous jumps of physical audio parameters, which creates audible artifacts. The second suggestion is the deliberate control of physical audio parameters that could be referred to as *psychoacoustic signal processing*. Physical audio parameters should either be used within ranges at which they affect exclusively one perceptual sound attribute. Or the undesired effect that one physical parameter has on a second perceptual attribute shall be counterbalanced by carefully adjusting other phys-

https://doi.org/10.21785/icad2019.018

ical audio parameters. Psychoacoustic signal processing treats the problem as a forward problem. However, it restricts the sonification designer to appropriate signal attributes and ranges. In that sense, [8] suggests a three-dimensional sonification approach that is based on timbre space in cylindrical coordinates. Elevation is represented by pitch, the radius by brightness, and discrete angles by certain musical instruments, i.e., timbre. Here, pitch is controlled by discrete notes played on the instrument and brightness by low pass filters. He points out the distinction between physical audio parameters and perceptual aspects of sound. However, he experienced that timbres in terms of instrument groups are nominally scaled (categorical), rather than rather than ordinal or even interval or ratio scaled. Hence, the angle in his approach is very vague and there exists no intuitive orthogonal or opposite direction.

In earlier studies we already presented an implementation of psychoacoustic signal processing for two-dimensional sonification [20] and some experimental validation with passive [21, 22] and interactive users [23, 24, 25] in a navigation task. In the paper at hand we modify the approach to enable three-dimensional sonification, e.g., for the sake of three-dimensional navigation. We will first explain the fundamentals of psychoacoustics and then its technical implementation in a sonification. Then, we discuss strengths and weaknesses of the approach. Finally, we give an outlook towards experimental validation, further development steps and potential application areas.

2. PSYCHOACOUSTIC SONIFICATION

For psychoacoustic sonification, principles of auditory perception are implemented in digital signal processing. The fundamentals are briefly summarized in this section. A deeper insight in psychoacoustics is given by [17]. Technical details of psychoacoustic signal processing can be found in [24, 25].

2.1. Psychoacoustics

The auditory system groups parts of incoming sounds so that the attributes of a group and their variations over time can be analyzed. This grouping is referred to as *auditory scene analysis* [26]. Complex tones are considered to exhibit at least five perceptual attributes, which are pitch, loudness, brightness, roughness, and fullness[27, ch. 32.2][28, 29]. Further attributes mentioned in the literature include subjective duration, tonalness and harmonicity [17, ch. 12 and pp. 363f][29]. All these perceived *auditory qualities* are nonlinear functions of more or less all *physical quantities*, which are the amplitude and the temporal and spectral distribution of frequencies. The physical attributes are physically independent, i.e., orthogonal, from one another. But they interfere perceptually. Likewise, the perceptual attributes are psychologically largely independent from one another, i.e., they can be regarded as orthogonal. But several physical attributes may affect them.

Auditory scene analysis: Auditory scene analysis is the psychological organization of sound and is closely related to Gestalt psychology [26, 20]. Portions of sound are integrated into *auditory streams* when they are in fair synchrony, have rather harmonic frequency relations and/or seem to come from the same spatial location. Streams are the auditory counterpart of visual objects. Streams have perceptual attributes like pitch, loudness and timbre. They are sustained over time as long as their components follow the law of continuity, proximity, common fate, timbre and closure, i.e., changes must be gradual and relations of the components must persist to some degree. Listeners can recognize some attributes of a second stream whilst consciously keeping track of another stream. To analyze details, the listener has to switch attention between streams. The presence or absence of a third stream can be noticed, but its details are not heard. Hence, sonification should be perceived as one auditory stream to ensure that all details are audible at once, without the need to switch attention [30]. Additional streams can be used to binary add simple pieces of information, like the plain presence or absence of a state or an item. When the specific task allows to concentrate on one stream at a time and switch attention if needed, two streams can be leveraged as well. Many researchers already highlighted the importance of auditory scene analysis principles and psychoacoustics in auditory display design [1, 5, 31, 8, 32].

Pitch: Perceived pitch is a multi-dimensional quality that consists of rectilinear *height* and circular *chroma*, which repeats every octave[33]. Pitch tends to be a rather linear function of fundamental frequency of harmonic complex tones. However, at fundamental frequencies above about 1 kHz the function becomes nonlinear. Sometimes, pitch is determined by signal period, e.g., in the case of a *missing fundamental*, or by the cutoff frequency, e.g., in the case of filtered peaked ripple noise [17, ch. 5]. Pitch can also be affected by signal amplitude, especially at very high or very low sound pressure levels. Pitch is neither binary nor instantaneous. A pitch strength exists, being generally higher for pure tones and complex tones compared to percussive, inharmonic, or noisy sounds. Pitch perception needs several milliseconds to build up.

Loudness: Loudness is closely related to signal amplitude [17, ch. 8][29, 34]. Increasing the amplitude or amplification gain makes a sound louder. However, different frequencies with equal amplitude tend to create different loudness sensations. Amplitude modulations slower than about 15 Hz are heard as loudness fluctuations, i.e., as *beats* [17, ch. 8 and 10][29].

Brightness: Auditory brightness mainly depends on the spectral distribution. It is considered the main contributor to timbre perception. The sensation of auditory brightness is closely related to auditory sharpness. The first is correlated with the spectral centroid [35]. The latter is explained by partial loudnesses along the Basilar membrane in the cochlea and considers masking effects[17, ch. 6][34, 29]. Shifting a spectral envelope towards higher frequencies makes a sound brighter. So does harmonic distortion, transposition towards higher frequencies, or applying a high-pass or shelving filter.

Roughness: Auditory roughness is considered another aspect of timbre [17, ch. 11][34, 29]. A rough sound is also referred to as being *jarring*, *harsh*, *raspy* or *blurred* [36, pp. 171, 349][27, ch. 32.2]. A pure tone sounds very smooth. Adding a second tone can have three effects that result from the critical bandwidth of the Basilar membrane, which is about 20% of a frequency. When its frequency deviates by more than 20% an interval may be heard, like a third or a fifth. An exception is tonal fusion, which may occur at frequency ratios of $1:2$, $1:3$, $1:4$ etc. Here, the tones may fuse and the additional tone creates the impression of changed timbre in terms of brightness, rather than the impression of an interval. When the frequency deviates by much less than the critical bandwidth, beating and a subtle pitch shift are perceived. Other frequency deviations up to 20% sound rough. The degree of roughness increases with an increasing number of non-beating frequencies within one critical bandwidth, as well as with the number

of critical bands that exhibit roughness. Roughness is easily created by means of amplitude modulations with frequencies between about 15 and 200 Hz, or by frequency modulations with frequencies around 50 Hz.

Many authors like [6, 32, 8] already identified pitch, loudness, sharpness, roughness and beating as potential parameters for psychoacoustic sonification.

Fullness: Fullness is sometimes referred to as *volume* or *sonority*[27, ch. 32.2][37, ch. 2]. Like brightness and roughness, it is an aspect of timbre. A full sound exhibits a broad frequency spectrum. The opposite is a *thin* or *narrow* sound, like that of a sinusoidal frequency or narrow band noise. Increasing the bandwidth, e.g., by means of distortion or frequency modulation, increases the degree of fullness. Decreasing the bandwidth by band pass filters lowers the degree of fullness.

Subjective Duration: Subjective duration only tends to be a linear function of physical duration in the range between 100 ms and several seconds [17, ch. 12]. For shorter sound events one has to divide signal duration by much more than 2 to half the subjective duration. When duration exceeds the capacity of the echoic memory, subjective timing becomes vague.

Further Attributes: Some studies consider *tonalness* and *harmonicity* as additional attributes of sounds [29, 34]. Tonalness ranges from tonal to noisy and depends on the number of frequency components and their amplitude and frequency relations. While a spectrum with discrete peaks sound tonal, a continuous frequency spectrum sounds noisy and loses pitch strength [17, ch. 5]. Pure and complex tones sound tonal and harmonic. The less peaks in a frequency spectrum resemble a harmonic series, i.e., $1:2:3:\ldots$, the more inharmonic it sounds. Inharmonic sounds can have one or multiple, more or less distinct, pitches. Spatial aspects of sound include source location and perceived source extent [38, 14]. The perceived source location is mainly a matter of the head-related transfer function, i.e., interaural level and time differences as well as spectral peaks and notches. These result from the sound propagation from the source to the ears, including deflections around and reflections from ears and torso. The ratio of direct sound to reverberation gives additional distance cues. Less is known about the perception of source extent. It seems to be affected by the coherence of ear signals and the presence of bass frequencies.

2.2. Sonification

In our previous two-dimensional sonification, chroma, loudness, and roughness were leveraged to communicate the direction and distance along two dimensions in Cartesian coordinates. Technical details of the implementation are provided in [25]. This sonification is the basis of our three-dimensional sonification that is described below. Table 1, summarizes the sonification parameters and their effect on the sound attributes, Fig. 1 helps for the understanding of the sound signal, Fig. 2 is a qualitative depiction of the mapping principle. Parameters are explained in the running text before their formulas are presented in Eqs. 1 to 9. Exemplary videos can be found on the first author's YouTube channel[1].

The sonification core is a harmonic *Shepard-tone* [33] with $N = 12$ partials. These partials act as carrier frequencies in terms

[1]See `https://tinyurl.com/ycwmdh8r` for previous 2D sonification and `http://tinyurl.com/y4um5odf` for the new 3D sonification.

direction	attribute	characteristic	function
left	pitch	falling speed	$\phi(\Delta x, t)$
right	pitch	rising speed	$\phi(\Delta x, t)$
up	beats	frequency	$g(\Delta y, t)$
down	fullness	degree	$\sigma(\Delta y)$
front	roughness	degree	$\beta(\Delta z)$
back	brightness	degree	$\mu(\Delta z)$

Table 1: Sonification principle and parameters when the target lies to the left/right/up/down/front/back.

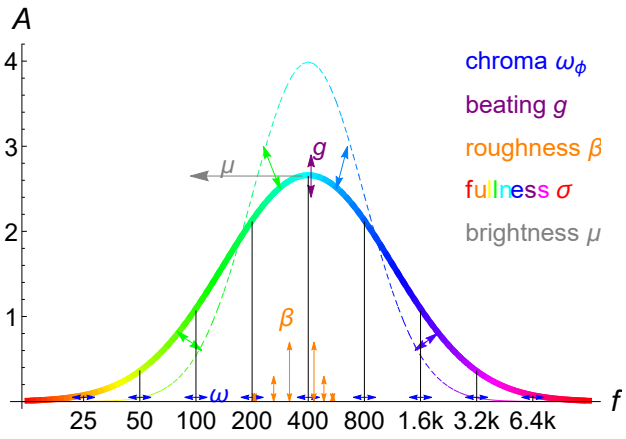

Figure 1: Sonification spectrum and parameters that modify the spectrum according to the direction and distance along the dimensions.

of additive frequency modulation synthesis. Their frequency ratios are $1:2^n$ with $n = 0,\ldots,N-1$, i.e., all frequencies are octaves of their neighbors. The amplitude $A(\phi_n(\Delta xt))$ depends on frequency f. On a logarithmic frequency scale the amplitude is a symmetric envelope that peaks in the center and is tapered off towards the sides. In Fig. 1 the envelope is represented by the colorful curve, the black vertical lines denote the carrier frequencies. The perceived pitch of Shepard tones exhibits only chroma but no height.

At the target x-coordinate the frequencies are steady. At all other x-coordinates the fundamental frequency and, accordingly, all partials move. When the target lies to the right, all frequencies rise. Many people perceive this as a rising pitch, the scientific expression is that chroma moves clockwise. In our implementation the lowest frequency is $f_0 = 3.125$ Hz, so the highest frequency will approach $f_{\max} < f_0 \times 2^{12} = 12,800$ Hz. When reaching this frequency the frequency will be shifted instantaneously back to f_0 Hz and start rising again. This way the Shepard tone creates the auditory illusion of an infinitely rising pitch while in fact it is a cyclic repetition. The envelope ensures that the lowest and highest frequencies become gradually (in)audible while the overall loudness is kept constant. While one frequency rises from f_0 to $f_{\max}$ the exact same spectrum repeats 11 times, i.e., each time a partial increased by one octave. The speed of rising determines the period of the repetition. The further the target x-coordinate lies to the right, the faster the frequencies rise, i.e., the shorter the repetition period and the higher the repetition frequency. To ensure a linear scaling, the repetition period should lie above the

lower limit of linear subjective duration, i.e., 100 ms. However, it is wise to restrict it to even longer periods, like 250 ms, i.e., 4 Hz, where the perceived fluctuation strength peaks [17, ch. 10]. This duration equals the mean syllabic length in speech, which lies between 150 and 300 ms; an order of magnitude that fairly corresponds to the integration time of 150 to 250 ms that has been found in the right non-primary auditory cortex[39].

A target to the left is denoted by decreasing frequencies accordingly, i.e., a descending pitch or counterclockwise chroma movement. Blue arrows in Fig. 1 indicate that frequencies move when the target is to the left or right.

The y-dimension is divided in two. When the target lies above, the envelope is periodically raised and reduced by a gain function $g(\Delta y, t)$. This is indicated by the purple arrow near the envelope peak in Fig. 1. This amplitude modulation is perceived as beating. The further above the higher the amplitude modulation frequency, i.e., the faster the beating. To ensure that the amplitude modulation does not create a roughness impression, the modulation frequency has to lie below 15 Hz. However, as for the pitch dimension, it is wise to keep modulation duration above the 100 ms threshold of linear duration perception and ideally even above the 150 ms integration time of the auditory system. When the target lies below, the envelope is deformed by a parameter $\sigma(\Delta y)$. The further below, the narrower the spectral bandwidth and the thinner the resulting sound. Reducing the spectral bandwidth has a drastic effect on perceived loudness. To balance out this effect, and keep loudness constant, the peak of the envelope is increased as the bandwidth decreases. Fig. 1 shows two exemplary values of σ, i.e., the thick envelope and the thin, dashed envelope. The deformation of the curve is indicated by colored arrows that connect the two curves.

The z-dimension is also divided in two. When the target lies in front, the sound becomes rough. The further to the front the rougher the sound. This is achieved by a frequency modulation of all carrier frequencies. A modulation frequency of $\omega_{\text{mod}} \approx 50$ Hz sounds rough for most carrier frequencies. The higher the modulation index $\beta(\Delta z)$, the rougher the sound. The frequency modulations not only create the impression of roughness, but also slightly increase loudness, create subtle inharmonicity and, at a extreme modulation depths, noisiness. Exemplary sidebands of one carrier frequency are illustrated as orange arrows in Fig. 1. When the target lies in the rear, the brightness is decreased, the further behind the target lies. This is achieved by a shifting the envelope towards lower frequencies by a function $\mu(\Delta z)$ illustrated in Fig. 1 by a gray arrow.

Fig. 2 illustrates how to navigate based on the sound attributes. The target lies in the center of the coordinate system. The sound tells the user where the target is. Accordingly, the symbols along the axes describe how the sound attributes change when moving along the dimensions. We refer to the current location of the user as *cursor*. When the cursor lies to the left of the target, the perceived pitch rises. This is indicated by blue angle brackets. The further to the left, the faster the pitch rises, indicated by the density of the angle brackets. Accordingly, when the cursor is far to the right of the target, the pitch will fall quickly. While approaching the target, the pitch changes more slowly. Finally, at the target x-coordinate the pitch is steady. When the cursor lies far below the target, a quick beating will be audible. While approaching the target the beating becomes slower. Finally, at the target height, loudness is steady, i.e., no beating can be heard. The beats are indicated by a fluctuating curve with a purple envelope that represents the beating frequency. When moving even further up, the

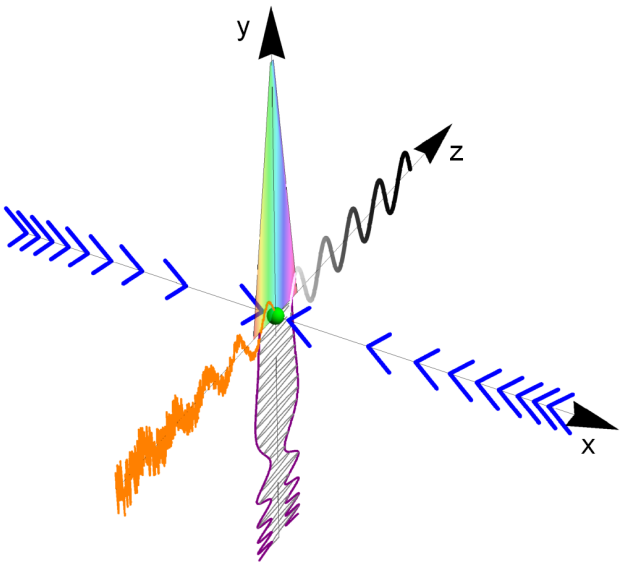

Figure 2: Navigation by sound attributes. The target lies at the origin of the coordinate system. The graphics symbolize how sound attributes change when moving along the corresponding axis.

fullness of the sound reduces more and more. This is indicated by a color spectrum that becomes narrower while the cursor goes up. When the cursor lies far behind the target, the sound is very rough. While approaching the target, roughness decreases, i.e., the sound becomes smoother. In the figure this is indicated by a jagged curve that becomes smoother towards the target. When the cursor lies far ahead of the target, the resulting sound is dull. While approaching the target the sound becomes brighter. This is indicated by the brightness level of the curve. One to three sound attributes can change at once, enabling three-dimensional navigation.

Additional elements that support navigation are added as segregated auditory streams. To enable navigation towards an extended target, a radius around the central target point is defined. The sonification guides towards the central target point. Pink noise is triggered as soon as the target region is reached. Pink noise is chosen because it is more subtle and pleasing than white noise, so it should not be perceived as a sudden alarm but as a calm confirmation sound in the background [25]. Slow beats are practically inaudible if a period takes seconds or even minutes. Likewise, there is not a specific point of maximum fullness that indicates the target y coordinate. Hence, a click is triggered as soon as the target height is reached. This click represents the x-z-plane in target-centered coordinate system. It is perceived as individual auditory stream because it is impulsive and neither belongs to the tonal Shepard tone nor to the noisy pink noise. Earlier studies showed that many novice users tend to trigger this click regularly to confirm that they are still at the target height [24]. Like fullness, roughness is gradual. To some extent, the degree of roughness is relative. A sound can be perceived as mildly rough, when heard after a very smooth sound. However, the same sound can be appear as perfectly smooth when heard directly after a very rough sound. Sometimes, it is difficult to identify the lowest possible degree of roughness, just as it may be difficult to identify the lowest audible pitch or loudness. Brightness also has no distinct minimum or

maximum. Due to the absence of a distinct point of origin, crossing the target z-coordinate is only heard because the sound that used to become fuller suddenly maintains its fullness but starts to become duller. This effect can be heard, but it is not very obvious. Hence, a short major chord is triggered every time the target z-coordinate is reached. This chord represents the x-y-plane. It confirms users that the target z-coordinate has been reached. Again, the chord is an individual auditory stream. Due to its short duration it sounds percussive, just as the click, but tonal. These three additional auditory streams only carry binary information. No attention switch is necessary to interpreted them. Note that the x-dimension is a ratio scale, because the steady pitch at $x = 0$ is an absolute zero. The same is true for the loudness fluctuation at $y = 0$. Roughness may also be considered as ratio scale, because a Shepard tone with octaves only contains no more than one frequency within each critical frequency band. However, brightness and fullness only represent an interval scale, because there is neither an obvious maximum of fullness at $y = 0$ nor of brightness at $z = 0$.

The sonification can be described by the formula

$$a_{\text{out}}(\Delta x, \Delta y, \Delta z, t) = g(\Delta y, t) \sum_{n=0}^{N} \Big[A(\phi_n(\Delta x, t), \Delta y, \Delta z) \times \cos\left[\omega_{\text{car}}(\phi_n(\Delta x, t))t + \beta(\Delta z)\cos(\omega_{\text{mod}} t)\right] \Big] , \quad (1)$$

where Δx, Δy and Δz describe the distance and direction between the current location and the designated target. The formulation is dynamic and real-time capable, so the current location and/or the target can move. It is explained in the same order as in Table 1 and the previous explanations.

The amplitude function

$$A(\phi_n(\Delta x, t), \sigma(\Delta y), \mu(\Delta z)) = \frac{e^{\frac{(\phi_n(\Delta x, t) - \mu(\Delta z))^2}{2\sigma^2(\Delta y)}}}{\sqrt{2\pi\sigma(\Delta y)}} \quad (2)$$

describes the symmetric envelope of the frequency spectrum. It is indirectly modified as a function of Δx, Δy and Δz. The phasor

$$\phi_n(\Delta x, t) = (\Delta x t + \varphi_n) \mod 1 \quad (3)$$

sweeps linearly from 0 to 1 at a frequency that depends on the distance along the x-axis. The larger the distance, the higher the sweep frequency. At negative Δx the sweep periodically decreases from 1 to 0. At a distance of $\Delta x = 0$ the sweep frequency is 0, so the phasor is a constant. In the equation mod is a modulo operation and φ_n is the initial phase of the nth carrier frequency. It is calculated as

$$\varphi_n = n/(N-1) . \quad (4)$$

Each phase has a corresponding frequency which is calculated as

$$f(\phi_n(\Delta x, t)) = f_0 2^{N\phi_n(\Delta x, t)} . \quad (5)$$

The envelope described in Eq. 2 is a Gaussian bell that peaks at the central frequency and tapers off towards the lower and upper frequencies. The phasor, Eq. 3, describes how frequencies move under this envelope. It is the only function of Δx. Perceptually, $\phi(\Delta x, t)$ controls the speed at which the pitch rises or falls.

In Eq. 1,

$$g(\Delta y, t) = \begin{cases} 1 + 0.5\sin(v_1 \Delta y t), & \text{if } \Delta y < 0 \\ 1, & \text{otherwise} \end{cases} \quad (6)$$

is an amplitude modulation. When the target lies above, it periodically modifies the gain of the signal. The modulation frequency is a function of the distance in y-direction. The further away the faster the modulation. The factor v_1 is scales the function to ensure a maximum beating frequency way below 15 Hz. Perceptually, Eq. 6 controls the speed of beating. The term $\sigma(\Delta y)$ in Eq. 2 is defined as

$$\sigma(\Delta y) = \begin{cases} \sigma_0 - v_2 \Delta y, & \text{if } \Delta y \geq 0 \\ \sigma_0, & \text{otherwise} \end{cases} . \quad (7)$$

Again, the term v_2 is just a scaling factor. The constant σ_0 is described later in relation to the brightness. The term $\beta(\Delta z)\cos(\omega_{\text{mod}} t)$ in Eq. 1 describes a nonlinear frequency modulation. The modulation index

$$\beta(\Delta z) = \begin{cases} a\Delta z^b + c, & \text{if } \Delta z < 0 \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

increases the further the target lies in the frontal direction. A higher modulation index increases the number and amplitudes of sidebands around each carrier frequency, i.e., new frequencies that are distributed symmetrically around the carrier frequencies. This is perceived as an increasing degree of roughness. A linear function and a power function are chosen because a linear increase starts extreme and then becomes subtle whereas a power function starts subtle but becomes extreme. Adding a constant c creates a sudden roughness jump at $\Delta z = 0$ which makes the target height better audible. The term

$$\mu(\Delta z) = \begin{cases} \mu_0 - \Delta z, & \text{if } \Delta z \geq 0 \\ \mu_0, & \text{otherwise} \end{cases} \quad (9)$$

in Eq. 2 shifts the envelope towards lower frequencies the further the target lies to the rear. This affects the brightness attribute. The further away the lower the brightness. The terms σ_0 and μ_0 have to be balanced carefully. The first has to be large enough to create a full sound at the target height. But it has to be small enough to taper off the signal towards the highest and lowest frequencies. This ensures that the instantaneous frequency shift from f_0 to f_{max} or vice versa creates no audible click due to the discontinuity. This is especially important at low values of μ, where the envelope is shifted towards the lower frequency end. If μ_0 is too low, the sonified range along the z-dimension is too small. If μ_0 is too large the target sound becomes too bright and shrill.

3. DISCUSSION

Our mapping was mainly driven by the need of three dimensions that are

- orthogonal in perception
- integrable (i.e., integrated into one auditory stream)
- linear
- continuous

and that exhibit

- a high resolution
- an audible origin of coordinates (without the need for a reference tone) .

This has been achieved by the described sonification principle. In addition, we tried to make it "sound worse" when the user moves in the wrong direction. On the x-axis the sound near the target feels like balancing. Pitch is slowly going up or down, like tuning of a guitar. However, when moving away from the target x-coordinate the pitch changes more and more rapidly. At the outer end this sounds like a siren that indicates danger. Beating works in a similar fashion. Near the target the beats are slow and gradual. But with increasing distance the beating becomes more hectic and at the outer end it sounds chopped, like an alarm, or the sound of car parking assistants near an obstacle. Roughness is known to be a contributor to auditory annoyance and a negative contributor to the sensory euphony of sound [40, 41]. So the further away, the rougher and less pleasing the sound. We successfully implemented these perceptual sound attributes in our two-dimensional sonification approach [42, 24].

The two new half-dimensions suggested in this paper are based on perceptual sound attributes that have been examined less comprehensively in the psychoacoustic literature. Here, we concentrated on physical, i.e., acoustical considerations. Very near sources tend to sound fuller and brighter compared to remote sources. This is due to near field effects and high-frequency attenuation in air. Low frequencies are not radiated from sound sources that are small compared to the wavelength. Instead, an acoustic short-circuit occurs and the low frequency energy stays in the near field of the source. The low frequencies are only audible in close proximity to the source. Furthermore remote sources sound more dull than nearer sources because heat exchange of short wavelengths in air attenuate high frequency energy. This effect becomes audible at distances in a range of tens to hundreds of meters. So we use one of these two physical effects along the brightness half-dimension and both along the fullness dimensions.

However, this consideration lacks psychoacoustic reasoning. Intuitively, we think that a narrow sound is less pleasing than a full sound. A narrow sound seems artificial, like through a telephone, or cheap loudspeakers. A full sound on the other hand is both warm and brilliant, which is desired at least in room acoustics [14, ch. 6]. But according to the literature a duller sound is more pleasing than a brighter sound [43, 34]. Here, it may be wise to flip the direction and make a sound increasingly bright when moving away from the target.

In our sonification up to three half-dimensions are heard at the same time. With the current mapping roughness and brightness manipulations cannot co-occur, because they occur at different directions on the same axis. The same is true for beats and fullness. In our two-dimensional approach we leveraged beats and roughness for the y-dimension. However, with our new sonification design the bandwidth parameter μ may not only affect fullness but also create beating sensation. This happens as soon as pitch moves and fullness is very low. Reducing the bandwidth to 1 to 4 frequencies the frequency-dependence of loudness sensation becomes obviously audible. As a result loudness fluctuates as a function of pitch. As a solution, we decided using beating and fullness as opposite directions of the same dimension. Now, a user knows that beats of a narrow sound belong to the fullness half-dimension and beats of a full sound belong to the beating dimension.

However, this solution may introduce another issue. Fullness and brightness can co-occur in this constellation. On broadband loudspeakers or headphones, this should not be a problem. The brightness half-dimension attenuates the highest frequencies only, while the low frequencies are kept. Furthermore, the shape of the envelope is kept. As a result loudness decreases together with brightness. The fullness half-dimension attenuates both the highest and lowest frequencies and increases the volume of the frequencies that are left. Here, the fullness does not affect the loudness. Unfortunately, budget loudspeakers may not be able to radiate low frequencies at all. In this case listeners cannot hear whether the lowest frequencies are attenuated or not, i.e., a listener can hardly tell the fullness and the brightness axis apart. In this special case the two half-dimensions perceptually interfere.

4. CONCLUSION

In this paper we discussed independent aspects of complex tones and how to leverage them for multi-dimensional sonification by means of psychoacoustic signal processing. Interpretability, orthogonality and linearity play a crucial role. In earlier works we have been able to derive, implement and examine a two-dimensional sonification that satisfied these criteria. In this paper, we demonstrated how this sonification can be modified and expanded to serve for three-dimensional sonification purposes. A strength of the sonification lies in the interaction. Users instantly hear when they move in the wrong direction and can correct their motion accordingly. Furthermore, the resolution of the sonification is scalable: the highest repetition rate of chroma cycles, beating, etc., can represent distance in the order or micrometers to kilometers. Such a scaling is not straightforward with spatial audio.

5. PROSPECTS

We are in the process of carrying out the same experiments with passive listeners as in [21, 20] to examine whether the half-dimensions are in fact orthogonal. Participants hear several sounds in a row, each representing one out of 16 fields on a map. With the 2D sonification 41% of the targets had been identified correctly, which is much better than chance (i.e., $1/16 \approx 6\%$).

For the new 3D sonification, each participant only evaluates two dimensions at a time, so the experiment can be kept short, the sonification easy to learn, and the results comparable to the earlier study. After some exploration of the system and experiments with 9 users, we decided to implement one of the solutions stated above: Brightness is now increased with increasing distance, to sound worse, i.e., shrill, at a large distance. The lowest degree of fullness is increased, so that even minimum fullness does not create loudness fluctuation. In a passive listening tests with the modified 3D sonification, users recognized over 55% of the target fields correctly. We currently analyze the experiment results and prepare a paper that contains the details of the sonification implementation, experimental conditions, and results.

After that, interactive experiments [25, 24, 23, 21] with the improved three-dimensional sonification will elicit whether the new half-dimensions are perceived as linear. This is a necessity to estimate the exact angle and distance of the target. Furthermore, exposing participants to all three dimensions will elicit whether the amount of information is reasonable or overwhelming for a user, and if a navigation task is manageable or overextending. Furthermore, only interactive experiments can show how comprehensible and effective the pink noise, the click and the major chord are. If interpretable, orthogonal, linear, and not overwhelming, the psychoacoustic three-dimensional sonification is ready for blind guidance in three-dimensional space and other multi-dimensional scenarios. Our team already started to add another sonification that

communicates the direction and distance of an obstacle. This task is demanding because this new sonification has to be integrated into one auditory stream which is segregated from the guidance sonification. Then, the user can focus on the guidance sonification but occasionally switch attention to the obstacle warning sonification that pops up every time an obstacle is closer than a predefined threshold. We are aware that such a six-dimensional sonification is ambitious and we assume a long learning phase. But the outcome is worth trying because such a sonification could communicate a large amount of data even in very complex scenarios and tasks.

We think the sonification has potential to act as an assistive tool in piloting, remote vehicle control, maneuvering and docking of spacecrafts, image-guided surgery, car parking and lane keeping, and as an audio game engine. Note that three orthogonal dimensions do not necessarily have to be spatial dimensions. The dimensions could also be heart rate, oxygen concentration and blood pressure in a patient monitoring task during anesthesia, the number of downloads, citations and mentions of a scientific book in a book metrics app, the average running speed, density and viewing direction of players in team sport modeling interventions or the charging status of gasoline, electricity, and coolant in a hybrid bus. For example our two-dimensional sonification has already been transferred successfully to a pulse-oximetry task [44]. With subtle modification the sonification can be adapted for movement analysis in dance and sports training and stroke rehabilitation, auditory graphs, auditory spirit level and many more. For continuous use over hours, the sonification is certainly too obtrusive and fatiguing. In such cases it it wise to switch it off when it is not needed.

6. ACKNOWLEDGMENT

We thank the students from the CURAT project at the University of Bremen who corrected an error in the sonification calculation and developed useful graphical user interfaces and a game-like environment for demonstrating and testing our sonification.

7. REFERENCES

[1] B. N. Walker and M. A. Nees, "Theory of sonification," in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Berlin: COST and Logos, 2011, ch. 2, pp. 9–39. [Online]. Available: http://sonification.de/handbook/

[2] J. Edworthy, S. Lexley, and I. Dennis, "Improving auditory warning design: Relationship between warning sound parameters and perceived urgency," *Human Factors*, vol. 33, no. 2, pp. 205–231, 1991. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/1860703

[3] M. Watson and P. M. Sanderson, "Intelligibility of sonifications for respiratory monitoring in anesthesia," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 45, no. 17, pp. 1293–1297, 2001. [Online]. Available: http://doi.org/10.1177/154193120104501708

[4] S. Barrass and G. Kramer, "Using sonification," *Multimedia Systems*, vol. 7, no. 1, pp. 23–31, 1999. [Online]. Available: http://doi.org/10.1007/s005300050108

[5] J. P. Bliss and R. D. Spain, "Sonification and reliability — implications for signal design," in *ICAD*, Montréal, Jun 2007, pp. 154–159. [Online]. Available: http://hdl.handle.net/1853/50028

[6] S. Ferguson, D. Cabrera, K. Beilharz, and H.-J. Song, "Using psychoacoustical models for information sonification," in *12th International Conference on Auditory Display*, London, Jun 2006. [Online]. Available: http://hdl.handle.net/1853/50694

[7] J. E. Anderson and P. Sanderson, "Sonification design for complex work domains: Dimensions and distractors," *Journal of Experimental Psychology: Applied*, vol. 15, no. 3, pp. 183–198, Mar 2009. [Online]. Available: http://doi.org/10.1037/a0016329

[8] S. Barrass, "A perceptual framework for the auditory display of scientific data," in *International Conference on Auditory Display*, Santa Fe, Nov 1994, pp. 131–145. [Online]. Available: http://hdl.handle.net/1853/50821

[9] D. Black, J. Hettig, M. Luz, C. Hansen, R. Kikinis, and H. Hahn, "Auditory feedback to support image-guided medical needle placement," *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, no. 9, pp. 1655–1663, 2017. [Online]. Available: http://doi.org/10.1007/s11548-017-1537-1

[10] F. Nagel, F.-R. Stöter, N. Degara, S. Balke, and D. Worrall, "Fast and accurate guidance – response times to navigational sounds," in *ICAD*, New York, NY, Jun 2014. [Online]. Available: http://hdl.handle.net/1853/52058

[11] A. Vasilijevic, K. Jambrosic, and Z. Vukic, "Teleoperated path following and trajectory tracking of unmanned vehicles using spatial auditory guidance system," *Applied Acoustics*, vol. 129, pp. 72–85, 2017. [Online]. Available: http://doi.org/10.1016/j.apacoust.2017.07.001

[12] T. Lokki and M. Gröhn, "Navigation with auditory cues in a virtual environment," *IEEE MultiMedia*, vol. 12, no. 2, pp. 80–86, April 2005. [Online]. Available: http://doi.org/10.1109/MMUL.2005.33

[13] J. Blauert, *Spatial Hearing. The Pychophysics of Human Sound Source Localization*, revised ed. Cambridge , MA: MIT Press, 1997.

[14] T. Ziemer, *Psychoacoustic Music Sound Field Synthesis: Creating Spaciousness for Composition, Performance, Acoustics, and Perception)*, ser. Current Research in Systematic Musicology. Cham: Springer, 2019, vol. 7.

[15] F. Rumsey, "Spatial audio. binaural challenges," *J. Audio Eng. Soc.*, vol. 42, no. 1, pp. 798–802, 2014.

[16] S. Bertet, J. Daniel, E. Parizet, and O. Warusfel, "Investigation on localisation accuracy for first and higher order ambisonics reproduced sound sources," *Acta Acustica united with Acustica*, vol. 99, no. 4, pp. 642–657, 2013. [Online]. Available: http://doi.org/10.3813/AAA.918643

[17] E. Zwicker and H. Fastl, *Psychoacoustics. Facts and Models*, second updated ed. Berlin, Heidelberg: Springer, 1999. [Online]. Available: http://doi.org/10.1007/978-3-662-09562-1

[18] F. Scheminzky, *Die Welt des Schalls*. Salzburg: Das Bergland-Buch, 1943.

[19] F. Pedrielli, E. Carletti, and C. Casazza, "Just noticeable differences of loudness and sharpness for earth moving machines," in *Proceedings of the European Conference on Noise Control 2008 (EURONOISE 2008)*, Paris, June 2008, pp. 1231–1236.

[20] T. Ziemer, D. Black, and H. Schultheis, "Psychoacoustic sonification for tracked medical instrument guidance," *Proceedings of Meetings on Acoustics*, vol. 30, 2017. [Online]. Available: http://doi.org/10.1121/2.0000557

[21] T. Ziemer, "Two-dimensional psychoacoustic sonification," in *33. Jahrestagung der deutschen Gesellschaft für Musikpsychologie (DGM)*, F. Olbertz, Ed., Hamburg, Sep 2017, pp. 60–61. [Online]. Available: https://www.researchgate.net/publication/319778727_Two-dimensional_psychoacoustic_sonification

[22] T. Ziemer and D. Black, "Psychoacoustically motivated sonification for surgeons," *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, no. 1, pp. 265–266, Jun 2017. [Online]. Available: http://doi.org/10.1007/s11548-017-1588-3

[23] T. Ziemer and H. Schultheis, "Perceptual auditory display for two-dimensional short-range navigation," in *Fortschritte der Akustik — DAGA 2018*. Munich: Deutsche Gesellschaft für Akustik, Mar. 2018, pp. 1094–1096.

[24] ——, "Psychoacoustic auditory display for navigation: an auditory assistance system for spatial orientation tasks," *J. Multimodal User Interfaces*, vol. Special Issue: Interactive Sonification, 2018. [Online]. Available: http://doi.org/10.1007/s12193-018-0282-2

[25] T. Ziemer, H. Schultheis, D. Black, and R. Kikinis, "Psychoacoustical interactive sonification for short range navigation," *Acta Acust. united Ac.*, vol. 104, no. 6, pp. 1075–1093, 2018. [Online]. Available: http://doi.org/10.3813/AAA.919273

[26] A. S. Bregman, *Auditory Scene Analysis*. Cambridge, MA: MIT Press, 1990.

[27] A. Schneider, "Perception of timbre and sound color," in *Springer Handbook of Systematic Musiwology*, R. Bader, Ed. Berlin, Heidelberg: Springer, 2018, ch. 32, pp. 687–726. [Online]. Available: http://doi.org/10.1007/978-3-662-55004-5_32

[28] W. Lichte, "Attributes of complex tones," *J. Exp. Psychol.*, vol. 28, pp. 455–480, 1941.

[29] T. Ziemer, Y. Yu, and S. Tang, "Using psychoacoustic models for sound analysis in music," in *Proceedings of the 8th Annual Meeting of the Forum on Information Retrieval Evaluation*, ser. FIRE '16, P. Majumder, M. Mitra, J. Sankhavara, and P. Mehta, Eds. New York, NY, USA: ACM, Dec 2016, pp. 1–7. [Online]. Available: http://doi.org/10.1145/3015157.3015158

[30] J. Anderson and P. Sanderson, "Designing sonification for effective attentional control in complex work domains," in *Proc. Human Factors and Ergonomins Society 48th annual meeting*, New Orleans, LA, Sep 2004. [Online]. Available: http://doi.org/10.1037/e577082012-006

[31] S. Barrass and V. Best, "Stream-based sonification diagrams," in *ICAD*, Paris, Jun 2008. [Online]. Available: http://hdl.handle.net/1853/49945

[32] G. Parseihian, C. Gondre, M. Aramaki, S. Ystad, and R. Kronland-Martinet, "Comparison and evaluation of sonification strategies for guidance tasks," *IEEE Trans. Multimedia*, vol. 18, no. 4, pp. 674–686, April 2016. [Online]. Available: http://doi.org/10.1109/TMM.2016.2531978

[33] R. N. Shepard, "Circularity in judgments of relative pitch," *The Journal of the Acoustical Society of America*, vol. 36, no. 12, pp. 2346–2353, 1964. [Online]. Available: http://doi.org/10.1121/1.1919362

[34] W. Aures, "Berechnungsverfahren für den sensorischen wohlklang beliebiger schallsignale (a model for calculating the sensory euphony of various sounds)," *Acustica*, vol. 59, no. 2, pp. 130–141, 1985.

[35] E. Schubert and J. Wolfe, "Does timbral brightness scale with frequency and spectral centroid?" *Acta Acustica united with Acustica*, vol. 92, no. 5, pp. 820–825, 2006. [Online]. Available: https://www.ingentaconnect.com/content/dav/aaua/2006/00000092/00000005/art00019

[36] H. von Helmholtz, *On the sensations of tone as a physiological basis for the theory of music*, 2nd ed. London: Longmans, Green, and Co., 1885.

[37] J. Meyer, *Acoustics and the Performance of Music. Manual for Acousticians, Audio Engineers, Musicians, Architects and Musical Instrument Makers*, 5th ed. Bergkirchen: Springer, 2009. [Online]. Available: http://doi.org/10.1007/978-0-387-09517-2

[38] T. Ziemer, "Source width in music production. methods in stereo, ambisonics, and wave field synthesis," in *Studies in Musical Acoustics and Psychoacoustics*, ser. Current Research in Systematic Musicoogy, A. Schneider, Ed. Cham: Springer, 2017, vol. 4, ch. 10, pp. 299–340. [Online]. Available: http://doi.org/10.1007/978-3-319-47292-8_10

[39] D. Poeppel, "The analysis of speech in different temporal-integration windows: cerebral lateralization asÔasymmetric sampling in time," *Speech Communication*, vol. 41, pp. 245–255, 2003.

[40] W. Ellermeier, A. Zeitler, and H. Fastl, "Predicting annoyance judgments from psychoacoustic metrics: Identifiable versus neutralized sounds," in *The 33rd International Congress and Exposition on Noise Control Engineering (inter-noise)*, Prague, Aug. 2004.

[41] W. Aures, "Ein Berechnungsverfahren der Rauhigkeit (a procedure for calculating auditory roughness)," *Acta Acust. united Ac.*, vol. 58, no. 5, pp. 268–281, 1985. [Online]. Available: https://www.ingentaconnect.com/content/dav/aaua/1985/00000058/00000005/art00005

[42] T. Ziemer and H. Schultheis, "A psychoacoustic auditory display for navigation," in *24th International Conference on Auditory Displays (ICAD2018)*, Houghton, MI, June 2018. [Online]. Available: http://doi.org/10.21785/icad2018.007

[43] B. Cardozo and R. van Lieshout, "Estimates of annoyance of sounds of different character," *Appl. Acoust.*, vol. 14, no. 5, pp. 323–329, 1981.

[44] S. Schwarz and T. Ziemer, "A psychoacoustic sound design for pulse oximetry," in *The 25th International Conference on Auditory Display (ICAD2019)*, Newcastle, June 2019.